Final Report for AOARD Grant FA2386-11-1-4106

Research Title: Developing decision-making skills using immersive VR

Date: 14/06/2013

Names of Principal Investigators:

1. **Name:** A/ Professor Gregor Kennedy
   Email: gek@unimelb.edu.au
   Institution: Centre for the Study of Higher Education
   Mailing Address: Level 1, 715 Swanston St, University of Melbourne, Victoria, 3010, Australia
   Phone: +61 3 8344 8937
   Fax: +61 3 8344 7576

2. **Name:** A/Professor James Bailey
   Email: baileyj@unimelb.edu.au
   Institution: Department of Computing & Information Systems
   Mailing Address: Room 7.09, Doug McDonell Building, The University of Melbourne, Victoria 3010, Australia.
   Phone: +61 3 8344 1319
   Fax: +61 3 9349 4596

3. **Name:** Professor Stephen O'Leary
   Email: sjoleary@unimelb.edu.au
   Institution: Department of Otolaryngology
   Mailing Address: Level 2, Royal Victorian Eye and Ear Hospital, 32, Gisborne St, East Melbourne, Victoria 3002, Australia
   Phone: +61 3 99298281
   Fax: +61 3 96631958

Period of Performance: 16/07/2012 – 14/06/2013

Abstract

The goal of this project was to investigate how decision-making skills in an immersive VR system could be improved using real-time feedback. We used a temporal bone surgical simulator as the teaching tool to train medical students on how to perform a cortical mastoidectomy. We used Random Forest based data mining models to assess the quality of the surgical technique and deliver timely feedback on how it can be improved. We performed an experiment with 24 medical students twelve of whom were given real-time feedback on surgical technique, and the remainder were not given any feedback. The test results suggest that the feedback delivered by the system not only had a high rate of accuracy, but was also effective in improving the surgical technique of medical students. Also, the responses of the participants to interview questions show that the system was highly usable and useful in learning surgical technique.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **19 MAR 2014**	2. REPORT TYPE **Final**	3. DATES COVERED **15-09-2011 to 29-03-2013**

4. TITLE AND SUBTITLE **Developing Decision-Making Skills Using Immersive Virtual Reality**	5a. CONTRACT NUMBER **FA23861114106**
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) **Gregor Kennedy**	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **University of Melbourne,Level 4, 766 Elizabeth Street,Melbourne 3010,Australia,AU,3010**	8. PERFORMING ORGANIZATION REPORT NUMBER **N/A**
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) **AOARD, UNIT 45002, APO, AP, 96338-5002**	10. SPONSOR/MONITOR’S ACRONYM(S) **AOARD**
	11. SPONSOR/MONITOR’S REPORT NUMBER(S) **AOARD-114106**

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
The goal of this project was to investigate how decision-making skills in an immersive VR system could be improved using real-time feedback. We used a temporal bone surgical simulator as the teaching tool to train medical students on how to perform a cortical mastoidectomy. We used Random Forest based data mining models to assess the quality of the surgical technique and deliver timely feedback on how it can be improved. We performed an experiment with 24 medical students twelve of whom were given real-time feedback on surgical technique, and the remainder were not given any feedback. The test results suggest that the feedback delivered by the system not only had a high rate of accuracy, but was also effective in improving the surgical technique of medical students. Also, the responses of the participants to interview questions show that the system was highly usable and useful in learning surgical technique.

15. SUBJECT TERMS
Human Attention & Performance, Virtual Reality, Training

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**9**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

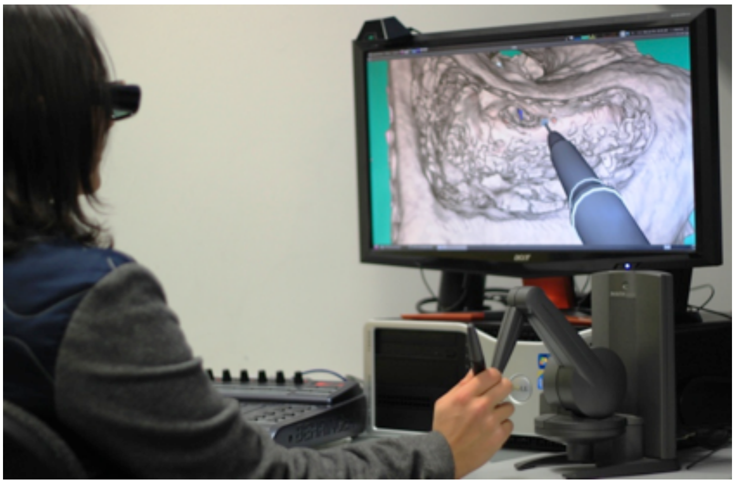
Introduction

Previous research suggests that existing professional training programs could be significantly improved with explicit cognitive skills training based on a model of deliberate practice [1]. Computer-based immersive virtual reality (VR) simulations have recently attracted attention as useful supplements to skills development training in a number of professions (e.g. aviation [2], defense [3], and health [4]).

When immersive VR simulations are used to support the training of novices typically an expert instructor provides feedback while tasks are undertaken by a trainee. An alternative instructional model is to use VR simulations as self-directed training tools with the VR system itself providing trainees with real-time feedback on their performance [5]. This project investigated how the metrics from a VR training environment could be harnessed to provide real-time feedback to trainees while undertaking deliberate practice.

Method

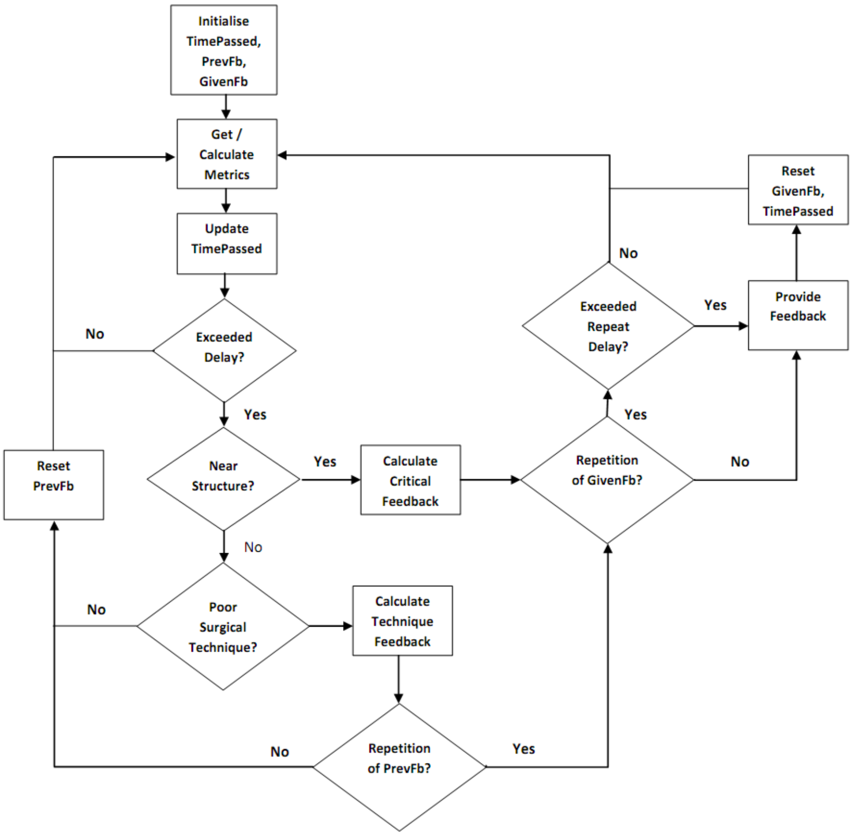
The first phase of the project involved the development of an automated feedback system for use with a VR training simulation, specifically in the area of ear surgery. The automated feedback system was developed to be used in concert with a temporal bone surgical simulator that has been developed by the University of Melbourne [6] and is shown in the figure below. This is a 3D immersive VR simulator that allows the user to interact with the simulation environment using a haptic device. The simulator can be used to perform surgical procedures such as cortical mastoidectomy, posterior tympanatomy, and cochleostomy. These procedures require the identification without injury of critical anatomical structures that are found within the temporal bone, including the nerve that animates the face, the major venous drainage from the head, the inner ear, and the dura.



The feedback system was developed based on models trained using data previously collected from expert and novice surgeons. The quality of the surgical technique and the optimal action for the user to undertake to bring his or her technique to the level of an expert was detected using a Random Forest data mining model and nearest neighbour techniques [7]. Feedback based on these models was generated in real time by analyzing a continuous data stream (at intervals of approximately 15Hz) generated by simulator.

Only if the same suggested feedback was proposed n times in a row did we deliver it to the user. This was done to increase the accuracy level of the feedback. In our trials of the system, we established that n=2 was the optimal number of repetitions required before providing feedback. Once a feedback was delivered, the system stopped processing the data for a t period of time. If the same feedback was repeated within a time period of T after the previously provided feedback, we ignored it. These delays were used to ensure that the user has time to correct their technique as suggested, and not to repeatedly bombard them with feedback. The delays we used in our experiment were t = 5s and T = 10s.

The following flow chart shows how the feedback system was designed.



Two different types of feedback were provided by the system:

- Suggestions on how to improve surgical technique if poor performance is detected.
- Warnings if the drill was near a critical anatomical structure.

Feedback was provided in the form of prerecorded audio advice on surgical technique. Participants could be given feedback in six areas and the feedback was to either increase or decrease one of the following stroke or system attributes:

- Stroke Length
- Stroke Speed
- Stroke Straightness
- Force
- Burr Size
- Zoom Level

Once the feedback system had been developed, the second phase of the project involved an experiment in which the effectiveness of the feedback system was assessed. Twenty-four students were recruited (13 MBBS, 10 MD, and 1 PhD) to participate in the experiment, all of whom had prior knowledge of the anatomy of the ear, but no surgical experience. All participants were shown a video tutorial of how to perform a cortical mastoidectomy, taught how to use the simulator, and after some time of familiarization, asked to perform this procedure on the simulator twice. Twelve participants were provided with real-time feedback from the feedback system described above, while the remaining twelve participants were not provided with feedback in this form.

The performance of all participants was recorded using a continuous data stream from the simulator and through the used of screen capture software for later analysis. At the end of the procedure, the participants were interviewed to obtain their views on the simulator in general and, in the case of the participants who received feedback, the feedback system in particular.

Results

The data obtained from the two groups of students were analyzed in different ways to evaluate three different aspects of the feedback system:

1. Effectiveness: Did the feedback provided assist students in improving their surgical technique?
2. Accuracy: How accurate was the given feedback when compared to that of an expert surgeon?
3. Usability: How usable did the students find the system, and was the feedback helpful to them?

Results in each of these areas are reported below.

Effectiveness

In order to evaluate whether the feedback provided was in fact effective in improving surgical technique, we compared the surgical behaviour of the two groups of students in terms of (i) analysis of surgical 'strokes' (ii) analysis of structure voxels drilled, and (iii) analysis of final bone shape.

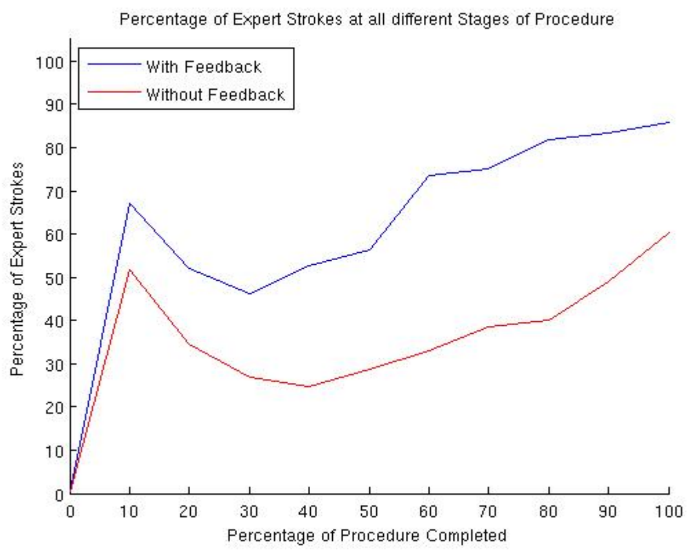
Analysis of Strokes
For each procedure in both participant groups, data streams associated with participants' strokes – the trajectory of the surgical drill – were extracted from the system and classified using the same Random Forest model used in the development of the feedback system. The percentage of expert strokes performed in each run by each participant was calculated. An ANOVA which compared both whether there were differences within each individual from Run 1 to Run 2 and whether there were differences between the two groups (feedback; no feedback) was performed. There was a significant between subjects effect ($F(22,1) = 29.06$; $p < .001$) and there was no significant within subjects effect ($F(22,1) = .02$; $p = .891$). These results indicate that there was a significant difference between groups with regards to stroke technique/expertise, but there was no difference in stroke technique or expertise for all participants between their first and second run.

Given the lack of difference in stroke technique from Run 1 to Run 2, the data for each participant across the two runs were combined (averaged) and then used in further analyses. Table 1 shows the results of an ANOVA test that shows a significant difference between the 'feedback' and 'non-feedback' groups with regards to average percentage of expert strokes recorded. It can be seen from Table 1 that there is a 58.5% increase in stroke expertise in the group that was provided with feedback with respect to the control group.

Percentage of Expert Strokes			
With Feedback	Without Feedback		
M (SD)	M (SD)	F	p
61.59 (16.19)	38.86 (13.11)	14.29	.001

Table 1: Analysis of expert stroke percentage for the two groups

The percentage of expert strokes for participants in each group during different stages of the procedure was also analysed. Figure 1 shows the results of this analysis at 10% intervals of completion.



Analysis of Structure Voxels Drilled

Damaging anatomical structures while performing surgical procedures could cause critical damage (facial paralysis, intracranial injury, severe haemorrhage or deafness), and as such, the aim is to expose the structures sufficiently to determine their location without damaging them. Therefore, the amount of damage caused to anatomical structures is an indication of expertise. As we provided warnings to one group of participants when they neared anatomical structures, it is also deemed a measure of the effectiveness of feedback. The percentage of voxels of anatomical structures removed by participants in each group were analysed using an ANOVA. Both within subjects (comparing participants' first and second run) and between subjects tests were not significant indicating no difference between groups or across individual participants' runs.

Analysis of Bone Shape

The shape of the virtual bones at the end of the procedure is another estimate of expertise, and is often used as a "summative" assessment in temporal bone dissection. In this analysis, the performance of participants was compared to that of expert surgeons (as had been established from the training data) to determine the likelihood that a bone had been drilled by an expert. ANOVA test results indicated that there were no significant differences within or between groups with regards to bone shape.

Accuracy

The accuracy of the feedback provided to participants by the system was determined through a post-experiment assessment carried out by an expert ear surgeon. The expert surgeon evaluated the feedback provided by the system for both runs performed by each participant. The accuracy of the system was assessed in three areas:

- When feedback was provided when stroke technique was acceptable (i.e. "False Positive Classifications")
- When participants' technique was accurately classified as "novice" but the content of the feedback was inaccurate (i.e. "Wrong Feedback")
- When feedback was not provided when stroke technique was unacceptable (i.e. "False Negative Classifications").

A total of 576 feedback messages were provided across the two runs of the twelve "feedback" participants. Of the feedback provided to participants:

- 39 feedback messages, or 6.7% of total feedback provided, were determined as "false positives",
- 52 feedback messages, or 9.0% of the total feedback provided, were assessed as "wrong feedback"; and
- 69, or 11.4% of the total feedback that would have been provided by an expert surgeon, were assessed as "false negatives".

An analysis of "Wrong Feedback" indicated that most of the inaccurate feedback (61.5%) related to incorrectly advising participants to alter the zoom level being applied in the simulator. Other areas of incorrect feedback related to stroke length (15.4%), stoke straightness (13.5%), stroke speed (5.8%), and the amount of force applied (3.9%).

Usability

The usability of the feedback system was assessed by analysing participants' answers to the following interview questions:

- Did you pay attention to the feedback and notice it while you completed the task?
- Did it assist you when you were completing the procedure or stages of it?
- Was it unhelpful, irrelevant or distracting when you were completing the procedure or stages of it?
- How could the provision of feedback by the system be improved or be made more useful?

The majority of the participants indicated that they noticed the feedback and that they paid attention to it when completing the task. Many also found it useful in completing the procedure. Participants commented particularly on the helpfulness of the warnings that were provided when they were close to a critical anatomical structure. For example, participant P06 stated: "it reminded me to be gentle near structures". Feedback on stroke technique was also deemed to be helpful. For example, participant P01 said: "particularly helpful was changing burr size and whether or not to zoom in". P07 said: "it gave me the confidence to go faster".

Only one participant indicated that the feedback was unhelpful while, a few found some of the feedback to be irrelevant, which is consistent with the errors that were detected in the feedback provided, which have been explained above. For example, P09 stated: "sometimes some of them weren't relevant, like when I was told to zoom out, I thought it was a good view already". A few students also mentioned that they were sometimes distracted by the feedback. For example, P01 said: "sometimes it was really out of the blue and caught you off guard".

Although the response to the simulator and the feedback system was overwhelmingly positive, there were some weaknesses the participants said they would like to see improved. The following is a list of the improvements that participants mentioned in the interviews:

- Restrict any contradictory feedback (e.g. 'Drill faster' when near a critical anatomical structure);
- Provide clearer feedback (e.g. feedback such as 'Use more curved strokes', 'You are being too tentative' were found to be too ambiguous);
- Provide more specific feedback (e.g. provide advice about what users should do when near a structure; indicate more specifically the direction in which to drill)
- Reduce the repetition of feedback (e.g. once a user has been told he or she is near a structure and has remained there for some time, reduce the number of times the warning is given);
- Provide greater assistance in the procedure (e.g. show which areas of the bone should be drilled; provide advice on when the end of a stage is achieved); and
- Provide visual feedback with additional information (e.g. indicate proximity to an anatomical structure; provide an ideal stroke path).

Discussion

The development of the feedback system and results of the preliminary trial provided in this report indicate that the feedback system performed exceptionally well with respect to effectiveness, accuracy, and usability. Participants who received feedback performed significantly better in terms of the expertise of their surgical technique (strokes) than participants who did not have access to the automated feedback system. While both groups of participants improved their performance across the procedure, a significant difference was maintained across the procedure between groups.

The feedback system also performed exceptionally well in terms of accuracy. In the provision of feedback, the classification of both false positives and false negatives was low (approximately 7% and 11% respectively). Moreover, the error rate in the content provided with the feedback was also low (9%). In our future work we will seek to have an increased number of experts rating the performance of the system to improve the reliability of this measure. We will also intend to integrate other data models into to the feedback system such as Pattern based models [8], and compare their performance with respect to the current Random Forest based model.

There were, however, no differences between participants who received feedback and those who did not in terms of percentage of structure voxels damaged. This may be because participants in the feedback condition received the proximity warnings too late to alter their technique, or it could even be that participants, as complete novices, were unable to expose critical anatomical structures without damaging them, despite the warnings. These are areas in which we will focus further investigation.

There was also no difference between the two groups in terms of the shape of the drilled bone at the end of the procedure. This is perhaps not surprising, as the feedback system did not explicitly provide advice on which areas of the bone to drill or what the end result of the drilling should look like. Provision of such location feedback is an avenue for future investigation.

An overwhelming majority of the participants found the feedback provided by the system to be useful. Participants reported few problems attending to the feedback; while at times some felt it was distracting. Participants also suggested ways of improving the feedback system, including reducing the provision of contradictory and ambiguous feedback and providing clearer and more specific feedback. Recommendations were also made about introducing different types of feedback and delivering them using different ways (e.g. visual overlays for areas to be drilled, visual quantitative feedback such as distance to anatomical structures). We intend to undertake further experimental work to consider the effectiveness of these modalities of feedback coming out of the feedback system.

In conclusion, the work undertaken in this project has led to the successful development of an automated feedback system that can be used alongside immersive simulation based environments. Though an initial research study, this system has been observed to perform extremely well in terms of effectiveness, accuracy, and usability. We are also confident that this is the first step in developing a system that successfully emulates the role of expert trainers in simulated training environments.

List of Publications

* Yun Zhou, James Bailey, Ioanna Ioannou, Sudanthi Wijewickrema, Gregor Kennedy, and Stephen O'Leary, *'Constructive Real Time Feedback for a Temporal Bone Simulator'*, In Proc. of International Conference on Medical Image Computing and Computer Assisted Intervention, 2013. Accepted.

* Yun Zhou, James Bailey, Ioanna Ioannou, Sudanthi Wijewickrema, Gregor Kennedy, and Stephen O'Leary, 'Pattern-Based Real-Time Feedback for a Temporal Bone Simulator', Proc. of the 19th ACM Symposium on Virtual Reality Software and Technology, 2013. Submitted.

* Sudanthi Wijewickrema, Ioanna Ioannou, and Gregor Kennedy, *'Adaptation of Marching Cubes for the Simulation of Material Removal from Segmented Volume Data'*, In Proc. of IEEE International Symposium on Computer-Based Medical Systems, 2013. Accepted.

References

[1] Ericsson, K. A., Krampe, R. T. et al. (1993). "The role of deliberate practice in the acquisition of expert performance." Psychological Review 100(3): 363-406.

[2] Howard, C. E. (2011). "Simulation & training: expecting the unexpected." Military & Aerospace Electronics 22(11): 12-23.

[3] Cosma, D. and Stanic, M.-P. (2011). "Implementing a Software Modeling-Simulation in Military Training." Revista Academiei Fortelor Terestre 16(2): 204-215.

[4] Hammoud, M. M., Nuthalapaty, F. S. et al. (2008). "To the point: medical education review of the role of simulators in surgical training." American Journal of Obstetrics & Gynecology 199(4): 338-343.

[5] Billings, D. R. (2012). "Efficacy of Adaptive Feedback Strategies in Simulation-Based Training." Military Psychology 24(2): 114-133.

[6] O'Leary, S., J., Hutchins, M. A., et al. (2008). "Validation of a Networked Virtual Reality Simulation of Temporal Bone Surgery." The Laryngoscope 118(6): 1040-1046.

[7] Zhou, Y., Bailey, J. et al. (2013)."Constructive Real Time Feedback for a Temporal Bone Simulator", In Proc. of International Conference on Medical Image Computing and Computer Assisted Intervention. Accepted.

[8] Zhou, Y., Bailey, J. et al. (2013)." Pattern-Based Real-Time Feedback for a Temporal Bone Simulator', Proc. of the 19th ACM Symposium on Virtual Reality Software and Technology, 2013. Submitted.

DD882: As a separate document, please complete and sign the inventions disclosure form.